Backward Pass

- **Please calculate $w3^+$**

  Please also explain how you get the formulas.

Step	Description	Formula
1	**Complete and optimize the formula to calculate w3's impact to E_{total}** **Expressing the formula with** ○ w3's impact to E_{o1} ○ w3's impact to E_{o2}	$\partial E_{total} / \partial w3 = \partial E_{o1} / \partial w_3 + \partial E_{o2} / \partial w_3$
2	**Complete and optimize the formula to calculate w3's impact to E_{o1}** **Expressing the formula with chain rule**	$\partial Eo1 / \partial w3 = (\partial net_{h2} / \partial w_3) * (\partial out_{h2} / \partial net_{h2}) * (\partial net_{o1} / \partial out_{h2}) * (\partial out_{o1} / \partial net_{o1}) * (\partial E_{o1} / \partial out_{o1})$
3	**Complete and optimize the formula to calculate w3's impact to E_{o2}** **Expressing the formula with chain rule**	$\partial Eo2 / \partial w3 = (\partial net_{h2} / \partial w_3) * (\partial out_{h2} / \partial net_{h2}) * (\partial net_{o2} / \partial out_{h2}) * (\partial out_{o2} / \partial net_{o2}) * (\partial E_{o2} / \partial out_{o2})$
4	**Complete and optimize the formula to calculate w3's impact to net_{h2}**	$\partial net_{h2} / \partial w3 = \partial(i_1 * w3 + i_2 * w4 + b_1 * 1) / \partial w3 = i_1 = 0.05$
5	**Complete and optimize the formula to calculate net_{h2}'s impact to out_{h2}**	$\partial out_{h2} / \partial net_{h2} = out_{h2} (1 - out_{h2}) = 0.24061341724$
6	**Complete and optimize the formula to calculate out_{h2}'s impact to net_{o1}**	$\partial net_{o1} / \partial out_{h2} = \partial(w_5 * out_{h1} + w_6 * out_{h2} + b_2 * 1) / \partial out_{h2} = w_6 = 0.45$
7	**Complete and optimize the formula to calculate net_{o1} to out_{o1}**	$\partial out_{o1} / \partial net_{o1} = out_{o1} (1 - out_{o1}) = 0.186815602$
8	**Complete and optimize the formula to calculate out_{o1} to E_{o1}** **Expressed with** $target_{o1}$ and out_{o1}	$\partial E_{o1} / \partial out_{o1} = \partial((target_{o1} - out_{o1})^2 / 2) / \partial out_{o1} = out_{o1} - target_{o1} = 0.74136507$
9	**$w3^+$** **New W3**	$W_3^+ = 0.24975114363$ $w_3 - n * \partial E_{total} / \partial w_3$

1.Complete and optimize the formula to calculate w_3's impact to E_{total}

$$\partial E_{total} / \partial w3 = \partial Eo1 / \partial w3 + \partial Eo2 / \partial w3$$

Hidden layer

A. Calculate the impact of w_3's impact to E_{o1}

Step	Formula
$\partial E_{o1} / \partial w3$	$(\partial net_{h2} / \partial w_3) * (\partial out_{h2} / \partial net_{h2}) * (\partial net_{o1} / \partial out_{h2}) * (\partial out_{o1} / \partial net_{o1}) * (\partial E_{o1} / \partial out_{o1})$
$\partial net_{h2} / \partial w3$	$\partial(i_1 * w_3 + i_2 * w_4 + b_1 * 1) / \partial w3 = i_1$
$\partial out_{h2} / \partial net_{h2}$	$out_{h2} (1 - out_{h2})$
$\partial net_{o1} / \partial out_{h2}$	$\partial(w_5 * out_{h1} + w_6 * out_{h2} + b_2 * 1) / \partial out_{h2} = w_6$
$\partial out_{o1} / \partial net_{o1}$	$out_{o1} (1 - out_{o1})$
$\partial E_{o1} / \partial out_{o1}$	$\partial((target_{o1} - out_{o1})^2 / 2) / \partial out_{o1} = out_{o1} - \text{target}_{o1}$

B. Calculate the impact of w_3's impact to E_{o2}

Step	Formula
$\partial E_{o2} / \partial w3$	$(\partial net_{h2} / \partial w_3) * (\partial out_{h2} / \partial net_{h2}) * (\partial net_{o2} / \partial out_{h2}) * (\partial out_{o2} / \partial net_{o2}) * (\partial E_{o2} / \partial out_{o2})$
$\partial net_{h2} / \partial w3$	$\partial(i_1 * w_3 + i_2 * w_4 + b_1 * 1) / \partial w_3 = i_1$
$\partial out_{h2} / \partial net_{h2}$	$out_{h2} (1 - out_{h2})$
$\partial net_{o2} / \partial out_{h2}$	$\partial(w_7 * out_{h1} + w_8 * out_{h2} + b_2 * 1) / \partial out_{h2} = w_8$
$\partial out_{o2} / \partial net_{o2}$	$out_{o2} (1 - out_{o2})$
$\partial E_{o2} / \partial out_{o2}$	$\partial((target_{o2} - out_{o2})^2 / 2) / \partial out_{o2} = out_{o2} - \text{target}_{o2}$

2. Complete and optimize the formula to calculate w3's impact to E_{o1}.

$$\partial Eo1 / \partial w3 = (\partial net_{h2} / \partial w_3) * (\partial out_{h2} / \partial net_{h2}) * (\partial net_{o1} / \partial out_{h2}) * (\partial out_{o1} / \partial net_{o1}) * (\partial E_{o1} / \partial out_{o1})$$

$\partial E_{total} / \partial w_3$ is read as “the partial derivative of E_{total} with respect to w_3“. You can also say “the gradient with respect to w_3“.

By applying the chain rule :

$$\partial E_{total} / \partial w_3 = \partial E_{total} / \partial out_{h2} * \partial out_{h2} / \partial net_{h2} * \partial net_{h2} / \partial w_3$$

Visually:

$$\partial E_{total} / \partial out_{h2} = \partial E_{01} / \partial out_{h2} + \partial E_{02} / \partial out_{h2}$$

- **Calculate $\partial E_{01} / \partial out_{h2}$:**

$\partial E_{01} / \partial out_{h2} = \partial E_{01} / \partial net_{01} * \partial net_{o1} / \partial out_{h2}$

But , $\partial E_{01} / \partial net_{01} = \partial E_{o1} / \partial out_{o1} * \partial out_{o1} / \partial net_{o1}$

Substitute $\partial E_{01} / \partial net_{01}$ **in** $\partial E_{01} / \partial out_{h2}$ **formula -**

$\partial E_{01} / \partial out_{h2} = \partial E_{o1} / \partial out_{o1} * \partial out_{o1} / \partial net_{o1} * \partial net_{o1} / \partial out_{h2}$

Substitute $\partial E_{01} / \partial out_{h2}$ **in** $\partial E_{total} / \partial w_3$ **formula for how w3's impact to** E_{o1} **-**

$\partial E_{total} / \partial w_3 = \partial E_{o1} / \partial out_{o1} * \partial out_{o1} / \partial net_{o1} * \partial net_{o1} / \partial out_{h2} * \partial out_{h2} / \partial net_{h2} * \partial net_{h2} / \partial w_3$

$\partial Eo1 / \partial w3 = (\partial net_{h2} / \partial w_3) * (\partial out_{h2} / \partial net_{h2}) * (\partial net_{o1} / \partial out_{h2}) * (\partial out_{o1} / \partial net_{o1}) * (\partial E_{o1} / \partial out_{o1})$

3. Complete and optimize the formula to calculate w3's impact to E_{o2}.

$\partial Eo2 / \partial w3 = (\partial net_{h2} / \partial w_3) * (\partial out_{h2} / \partial net_{h2}) * (\partial net_{o2} / \partial out_{h2}) * (\partial out_{o2} / \partial net_{o2}) * (\partial E_{o2} / \partial out_{o2})$

$\partial E_{total} / \partial w_3 = \partial E_{total} / \partial out_{h2} * \partial out_{h2} / \partial net_{h2} * \partial net_{h2} / \partial w_3$

Visually:

$\partial E_{total} / \partial out_{h2} = \partial E_{01} / \partial out_{h2} + \partial E_{02} / \partial out_{h2}$

- **Calculate $\partial E_{02} / \partial out_{h2}$:**

$\partial E_{02} / \partial out_{h2} = \partial E_{02} / \partial net_{02} * \partial net_{o2} / \partial out_{h2}$

But, $\partial E_{02} / \partial net_{02} = \partial E_{o2} / \partial out_{o2} * \partial out_{o2} / \partial net_{o2}$

Substitute $\partial E_{02} / \partial net_{02}$ **in** $\partial E_{02} / \partial out_{h2}$ **formula -**

$\partial E_{02} / \partial out_{h2} = \partial E_{o2} / \partial out_{o2} * \partial out_{o2} / \partial net_{o2} * \partial net_{o2} / \partial out_{h2}$

Substitute $\partial E_{02} / \partial out_{h2}$ **in** $\partial E_{total} / \partial w_3$ **formula for how w3's impact to** E_{o2} **-**

$\partial E_{total} / \partial w_3 = \partial E_{o2} / \partial out_{o2} * \partial out_{o2} / \partial net_{o2} * \partial net_{o2} / \partial out_{h2} * \partial out_{h2} / \partial net_{h2} * \partial net_{h2} / \partial w_3$

$\partial E_{o2} / \partial w_3 = (\partial net_{h2} / \partial w_3) * (\partial out_{h2} / \partial net_{h2}) * (\partial net_{o2} / \partial out_{h2}) * (\partial out_{o2} / \partial net_{o2}) * (\partial E_{o2} / \partial out_{o2})$

4.Complete and optimize the formula to calculate w3's impact to net_{h2}.

how much does the total net input of h_2 change with respect to w_3.

$$\partial net_{h2} / \partial w3 = \partial(i_1 * w3 + i_2 * w4 + b_1 * 1) / \partial w3$$

$$\partial(i_1 * w3 + 0 + 0) / \partial w3 = i_1 = 0.05$$

5. Complete and optimize the formula to calculate net_{h2}'s impact to out_{h2}.

how much does the output of **h_2** change with respect to its total net input :

The partial derivative of Logistic Function is the output multiplied by 1 minus the output:

$$\partial out_{h2} / \partial net_{h2} = out_{h2} (1 - out_{h2})$$

$$= 0.59688437826(1 - 0.59688437826) = 0.24061341724$$

6. Complete and optimize the formula to calculate out_{h2}'s impact to net_{o1}.

how much does the total net input of o_1 change with respect to output of **h_2** :

$$\partial net_{o1} / \partial out_{h2} = \partial(w_5 * out_{h1} + w_6 * out_{h2} + b_2 * 1) / \partial out_{h2}$$

$$\partial(0 + w_6 * out_{h2} + 0) / \partial out_{h2} = w_6 = 0.45$$

7. Complete and optimize the formula to calculate net_{o1} to out_{o1}.

how much does the output of o_1 change with respect to its total net input?

The partial derivative of Logistic Function is the output multiplied by 1 minus the output:

$$\partial out_{o1} / \partial net_{o1} = out_{o1} (1 - out_{o1})$$

$$\frac{\partial out_{o1}}{\partial net_{o1}} = out_{o1}(1 - out_{o1}) = 0.75136507(1 - 0.75136507) = 0.186815602$$

8. Complete and optimize the formula to calculate out_{o1} to E_{o1}.

how much does the total error change with respect to the output?

$$E_{total} = \frac{1}{2}(target_{o1} - out_{o1})^2 + \frac{1}{2}(target_{o2} - out_{o2})^2$$

$$\frac{\partial E_{total}}{\partial out_{o1}} = 2 * \frac{1}{2}(target_{o1} - out_{o1})^{2-1} * -1 + 0$$

$$\partial E_{o1} / \partial out_{o1} = \partial((target_{o1} - out_{o1})^2 / 2) / \partial out_{o1}$$

$-(target - out)$ is sometimes expressed as $out - target$

$$= out_{o1} - target_{o1}$$

$$\frac{\partial E_{total}}{\partial out_{o1}} = -(target_{o1} - out_{o1}) = -(0.01 - 0.75136507) = 0.74136507$$

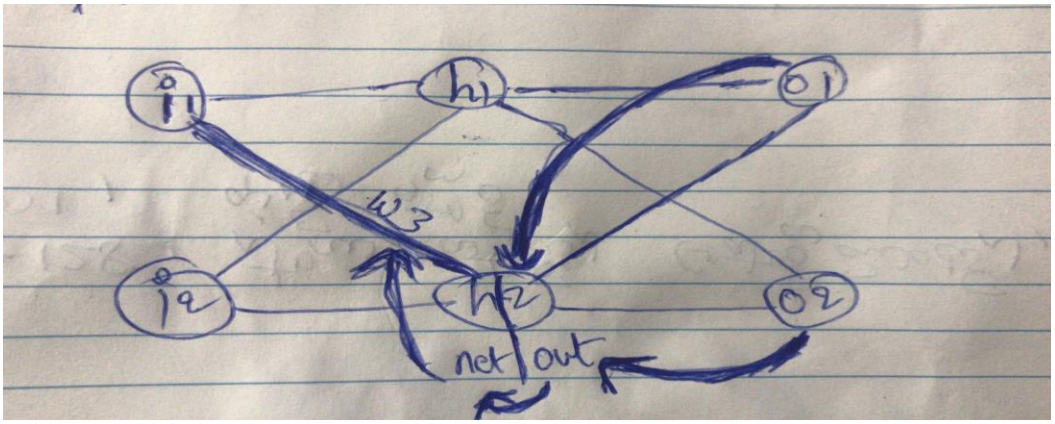
9. $W3^+$ (Caluclated $W3^+$ value below) :

$W_3^+ = w_3 - n * \partial Etotal / \partial w_3$ ==>

$0.25 - (0.5 * 0.00049771273)$ ==> 0.24975114363

$W_3^+ = 0.24975114363$

We'll continue the backwards pass by calculating new values for w_3.



$\partial E_{total} / \partial w3 = \partial E_{total} / \partial out_{h2} * \partial out_{h2} / \partial net_{h2} * \partial net_{h2} / \partial w_3$

$E_{total} = E_{01} + \partial E_{02}$

Visually:

$\partial E_{total} / \partial out_{h2} = \partial E_{01} / \partial out_{h2} + \partial E_{02} / \partial out_{h2}$

We need to figure out each piece in this equation.

$$net_{h1} = w_1 * i_1 + w_2 * i_2 + b_1 * 1$$

$$net_{h1} = 0.15 * 0.05 + 0.2 * 0.1 + 0.35 * 1 = 0.3775$$

We then squash it using the logistic function to get the output of h_1:

$$out_{h1} = \frac{1}{1+e^{-net_{h1}}} = \frac{1}{1+e^{-0.3775}} = 0.593269992$$

$$net_{h2} = w_3 * i_1 + w_4 * i_2 + b1 * 1$$

$$= 0.25 * 0.05 + 0.30 * 0.1 + 0.35 * 1 = 0.3925$$

$$out_{h2} = 1/1+e^{-net_{h2}} = 1/1+e^{-0.3925} = 0.59688437826$$

$$net_{o1} = w_5 * out_{h1} + w_6 * out_{h2} + b_2 * 1$$

$$net_{o1} = 0.4 * 0.593269992 + 0.45 * 0.596884378 + 0.6 * 1 = 1.105905967$$

$$out_{o1} = \frac{1}{1+e^{-net_{o1}}} = \frac{1}{1+e^{-1.105905967}} = 0.75136507$$

$$E_{total} = \frac{1}{2}(target_{o1} - out_{o1})^2 + \frac{1}{2}(target_{o2} - out_{o2})^2$$

$$\frac{\partial E_{total}}{\partial out_{o1}} = 2 * \frac{1}{2}(target_{o1} - out_{o1})^{2-1} * -1 + 0$$

$$\frac{\partial E_{total}}{\partial out_{o1}} = -(target_{o1} - out_{o1}) = -(0.01 - 0.75136507) = 0.74136507$$

$$out_{o1} = \frac{1}{1+e^{-net_{o1}}}$$

$$\frac{\partial out_{o1}}{\partial net_{o1}} = out_{o1}(1 - out_{o1}) = 0.75136507(1 - 0.75136507) = 0.186815602$$

A) We know that out_{h2} affects both out_{o1} and out_{o2} therefore the ∂E_{total} ∂out_{h2} needs to take into consideration its effect on the both output neurons:

$$\partial E_{total} / \partial out_{h2} = \partial E_{01}/\partial out_{h2} + \partial E_{02}/\partial out_{h2}$$

1. Starting with $\partial E_{01}/\partial out_{h2}$:

$$\partial E_{01}/\partial out_{h2} = \partial E_{01}/\partial net_{01} * \partial net_{01}/\partial out_{h2}$$

a. We can calculate $\frac{\partial E_{o1}}{\partial net_{o1}}$ using values, we calculated earlier:

$$\partial E_{01}/\partial net_{01} = \partial E_{01}/\partial out_{01} * \partial out_{01}/\partial net_{01}$$

$$\frac{\partial E_{o1}}{\partial net_{o1}} = \frac{\partial E_{o1}}{\partial out_{o1}} * \frac{\partial out_{o1}}{\partial net_{o1}} = 0.74136507 * 0.186815602 = 0.138498562$$

b. And $\partial net_{01}/\partial out_{h2}$ is equal to w_6:

$$net_{o1} = w_5 * out_{h1} + w_6 * out_{h2} + b_2 * 1$$

$$\partial net_{01}/\partial out_{h2} = 0.45$$

c. Plugging them in:

$$\partial E_{01}/\partial out_{h2} = \partial E_{01}/\partial net_{01} * \partial net_{01}/\partial out_{h2}$$

$$= 0.138498562 * 0.45 = 0.0623243529$$

2. Following the same process for $\partial E_{02}/\partial out_{02}$, we get:

$$\partial E_{02}/\partial out_{h2} = \partial E_{02}/\partial net_{02} * \partial net_{02}/\partial out_{h2}$$

a. We can calculate $\partial E_{02}/\partial net_{02}$ using values, we calculated earlier:

$$\partial E_{02}/\partial net_{02} = \partial E_{02}/\partial out_{02} * \partial out_{02}/\partial net_{02}$$

$$net_{h2} = w_3 * i_1 + w_4 * i_2 + b1 * 1$$

$= 0.25 * 0.05 + 0.30 * 0.1 + 0.35 * 1 = 0.3925$

$out_{h2} = 1/1+e^{-net_{h2}} = 1/1+e^{-0.3925} = 0.59688437826$

$$\mathbf{net_{02} = \ w_7 * out_{h1} + w_8 * out_{h2} + b_2 * 1}$$

$$\mathbf{= 0.50 * 0.593269992 + 0.55 * 0.59688437826 + 0.60 * 1 ==> 1.22492140404}$$

$$\mathbf{out_{02} = 1/1+e^{-net_{02}} = 1/1+e^{-1.22492140404} ==> 0.77292846531}$$

$$E_{total} = \frac{1}{2}(target_{o1} - out_{o1})^2 + \frac{1}{2}(target_{o2} - out_{o2})^2$$

- When we take the partial derivative of the total error with respect to out_{02}, the quantity ½$(target_{01}-out_{01})$^2 becomes zero because out_{02} does not affect it which means we're taking the derivative of a constant which is zero.

$$\mathbf{\partial E_{total}/\partial out_{02} = 0 + 2 * \tfrac{1}{2}(target_{02}-out_{02})^{2-1} * -1}$$

$$\mathbf{\partial E_{total}/\partial out_{02} = -(target_{02}-out_{02}) = -(0.99 - 0.77292846531) = -0.21707153469}$$

$$\mathbf{\partial out_{02}/\partial net_{02} = out_{02}\ (1- out_{02})}$$

$$\mathbf{= 0.77292846531(1- 0.77292846531) = 0.17551005282}$$

$$\mathbf{\partial E_{02}/\partial net_{02} = \partial E_{02}/\partial out_{02} * \partial out_{02}/\partial net_{02}}$$

$$\mathbf{= - 0.21707153469 * 0.17551005282 = -0.03809823651}$$

b. And $\mathbf{\partial net_{02}/\partial out_{h2}}$ is equal to $\mathbf{w_8}$:

$$\mathbf{Net_{02}} = w_7 * out_{h1} + w_8 * out_{h2} + b_2 * 1$$

$$\mathbf{\partial net_{02}/\partial out_{h2} = 0.55}$$

c.plugging them in:

$$\mathbf{\partial E_{02}/\partial out_{h2} = \partial E_{02}/\partial net_{02} * \partial net_{02}/\partial out_{h2}}$$

$$\mathbf{= -0.03809823651 * 0.55 = -0.02095403008}$$

3. Therefore:

$$\mathbf{\partial \mathit{E_{total}} / \partial out_{h2} = \ \partial E_{01}/\partial out_{h2} + \partial E_{02}/\partial out_{h2}}$$

$$\mathbf{\partial \mathit{E_{total}} / \partial out_{h2} = 0.0623243529 + -0.02095403008 = 0.04137032282}$$

B . Now that we have $\partial E_{total} / \partial out_{h2}$, we need to figure out $\partial out_{h2} / \partial net_{h2}$ and then $\partial net_{h2} / \partial w_3$ for each weight:

$$\mathbf{out_{h2} = 1/1+e^{-net_{h2}} = 1/1+e^{-0.3925} = 0.59688437826}$$

$$\mathbf{\partial \mathit{out}_{h2} / \partial \mathit{net}_{h2} = \mathit{out}_{h2}\ (1 - \mathit{out}_{h2}) = 0.24061341724}$$

$$0.59688437826(1- 0.59688437826) = 0.24061341724$$

C. We calculate the partial derivative of the total net input to h_2 with respect to w_3 the same as we did for the output neuron:

$$\mathbf{net_{h2}} = i_1 * w_3 + i_2 * w_4 + b1 * 1$$

$$\mathbf{net_{h2} / \partial w_3 = i_1 = 0.05}$$

D. Putting it all together

$$\partial E_{total} / \partial w3 = \partial E_{total} / \partial out_{h2} * \partial out_{h2} / \partial net_{h2} * \partial net_{h2} / \partial w_3$$

$$\partial E_{total} / \partial w_3 = 0.04137032282 * 0.24061341724 * 0.05 \implies 0.00049771273$$

E. Find the new weight

To decrease the error, we then subtract this value from the current weight (optionally multiplied by some learning rate, eta, which we'll set to 0.5):

$$W_3^+ = w_3 - n * \partial E_{total} / \partial w_3$$

$$0.25 - (0.5 * 0.00049771273) \implies 0.24975114363$$

$$W_3^+ = 0.24975114363$$